Il lessico di frequenza dell'italiano parlato e la didattica dell'italiano

Emilio D'Agostino

Abstract

The linguistic-historical evolution of the Italian community has been well described in its general lines and fundamental mechanisms. In the last decades the process of linguistic unification in Italy has been strengthened; which is confirmed by data on the diffusion of italophony and the regression of exclusive dialectophony.

In relation to this De Mauro has sustained that, nowadays, in the Italian community about 34% of the population only speaks Italian, 23% usually speaks only one of the dialects, while the majority uses Italian for speaking and most especially for writing, while at the same time adopting one dialect for speaking.

Sociolinguistic studies the tendency towards the consolidation of a standard Italian in which new features become established while others disappear. In particular, characteristics originally inherent in speech seem to have been strengthening in writing. The publishing of LIP (Lessico di Frequenza dell'Italiano Parlato) appears to be an interesting document and a valid tool of research to better follow the processes mentioned above and investigate thoroughly the status of the spoken language.

This work represents a first attempts to draw the attention of linguistic pedagogy on some mechanisms in contemporary Italian, as they can be defined or redefined in view of the information provided by the LIP corpus and of the problems related to it.

1. L'esperienza del Lessico di Frequenza dell'Italiano Parlato

L'evoluzione storico-linguistica della collettività italiana è stata ben descritta,[1] nelle sue linee generali e nei suoi meccanismi fondamentali. I processi di unificazione e omogeneizzazione linguistici dell'Italia si sono andati rafforzando negli ultimi decenni e i dati sulla diffusione dell'italofonia e il regresso della dialettofonia esclusiva lo confermano. De Mauro, a tale riguardo, ha avuto più volte modo di sostenere come oggi, nella collettività italiana, circa il 34% della popolazione parli esclusivamente italiano, il 23% parli abitualmente soltanto uno dei dialetti, mentre la maggioranza parla e, soprattutto, scrive italiano e, allo stesso tempo, parla

1. Basti qui ricordare De Mauro (1963) cui ha fatto seguito tutta quella considerevole mole di lavori che ad esso si sono ispirati e, parimenti, menzionare come utile lavoro di sintesi aggiornata i due volumi curati da Sobrero (1993).

VISIT…

un dialetto. Le indicazioni provenienti dalla maggior parte della letteratura sociolinguistica sottolineano la tendenza (almeno) al consolidamento di un nuovo standard dell'italiano[2] nel quale si affermano nuovi tratti, mentre altri scompaiono. In particolare, nell'uso scritto si sarebbero andate consolidando caratteristiche in origine proprie del parlato. Per la morfosintassi, ad esempio, ne sarebbero prova fenomeni come, in primo luogo, la semplificazione dei sistemi pronominali, tradizionalmente considerati complessi e troppo differenziati; fenomeni come le dislocazioni a sinistra e a destra e la frase scissa o l'uso dell'ormai leggendario «*che* polivalente». Essi costituirebbero prove della riconfigurazione del sistema. Si assisterebbe altresì ad una tendenza verso una semplificazione dei costrutti, con una spinta all'uso di costrutti limitati, ma molto frequenti, e verso una preferenza della costruzione paratattica rispetto a quella ipotattica. Verso la modificazione, infine, dell'economia delle nominalizzazioni e la modificazione parziale dei valori di categorie grammaticali come quelle aspettuali. Ma il dato di fatto più significativo sarebbe comunque un avvicinamento tra i caratteri dello scritto e quelli del parlato e, più in particolare, l'affermarsi di un uso parlato dell'italiano, anche informale, vieppiù lontano da quello dominato, come lo è stato a lungo, dalla forte influenza dialettale e locale.

La pubblicazione del *Lessico di Frequenza dell'Italiano Parlato*[3] (LIP) offre uno strumento di documentazione e di indagine assai stimolante, da un lato, per poter seguire, dati sperimentali alla mano, i processi cui abbiamo accennato e, dall'altro, per poter meglio approfondire le conoscenze sullo statuto della lingua parlata.

Il nostro contributo in questa sede rappresenta un tentativo — a dire il vero più che sintetico — di richiamare l'attenzione della pedagogia linguistica su alcuni meccanismi dell'italiano contemporaneo, così come è possibile definirli o ridefinirli alla luce delle indicazione del corpus del LIP e delle problematiche a questo collegate.

1.1. I lessici di frequenza dell'italiano

Quando nel 1972 apparve il *Lessico di frequenza della lingua italiana contemporanea* (LIF), poche erano state le esperienze di linguistica computazionale applicata alla lingua italiana, se si escludono soprattutto i lavori definiti «pionieristici» di R. Busa nel campo degli spogli elettronici e i contributi dello stesso A. Zampolli e dell'équipe del CNUCE di Pisa. In particolare, nel dominio dell'elaborazione dei lessici di frequenza, gli stessi precedenti che si possono ricordare, e cioè Thompson (1927), Knease (1933), Skinner (1935) e Migliorini (1943), erano tutti risultato di spogli manuali. Accomunati fondamentalmente da finalità didattiche, i vari lessici

2. Sulla questione del cosiddetto «neostandard» si veda, oltre al già citato Sobrero (1993), Berruto (1987) e Berretta (1994).
3. Cfr. De Mauro et alii (1993).

di frequenza citati erano anche caratterizzati dalla limitatezza dei vari *corpora* sui quali erano sviluppate le relative liste di frequenza. In particolare, Thompson (risultato di una tesi di dottorato) analizzava un campione di 100000 parole prese da manuali di lettura d'italiano pubblicati negli USA, la lista elaborata era costituita da 500 parole; Knease, anch'esso nato come tesi, rappresentava invece uno spoglio manuale di circa 4000 parole tratte da fonti esclusivamente letterarie tra '800 e '900; Skinner individuava una lista di 3000 parole prese da materiali didattici adottati negli USA; Migliorini, infine, era sviluppato unicamente sulla base del sapere linguistico dell'autore, per scopi didattici, e forniva una lista di 1500 parole ritenute fondamentali.

Il LIF nasce, invece, con dimensioni e finalità del tutto diverse. Innanzitutto, esso è costituito da uno spoglio di 500000 parole della lingua italiana contemporanea nell'uso scritto e porta all'individuazione di 5000 lemmi. La principale caratteristica del LIF sta nel fatto che i testi che costituiscono *corpus* sono contraddistinti da una relativamente forte sincronicità e attualità. Infatti partono dal 1945 e vanno fino al 1968, e sono organizzati in cinque gruppi: teatro, romanzi, cinema, periodici, sussidiari. In secondo luogo, nella sua definizione vengono ben individuate le problematiche di fondo della compilazione dei dizionari di frequenza. Nell'*Introduzione* si può leggere:

> I testi sottoposti a spoglio si configurano nelle intenzioni dei compilatori dei dizionari di frequenza, come campioni della lingua intesa quale universo statistico e di conseguenza pongono i problemi della campionatura per i quali la statistica ha elaborato tutta una serie di metodologie e di tecniche. […] Da un lato occorre delimitare l'universo, dall'altro assicurarsi della rappresentatività del campione. Se si conoscono bene le caratteristiche dell'universo, si può fabbricare un campione in cui ciascuna di queste caratteristiche sia distribuita secondo le stesse proporzioni, con una sottile stratificazione […] Che non sia questo il caso della lingua è facilmente intuibile. Le frequenze si osservano in un testo, cioè in una realizzazione del sistema linguistico da parte di un individuo, realizzazione determinata da un lato dalle sue caratteristiche personali, dall'altro da quelle della comunità cui appartiene, caratterizzata sul piano sociale, geografico, storico, ecc. Se il testo è un testo letterario, nella sua produzione influiscono anche le regole e le peculiarità del genere letterario e in ogni caso la situazione in cui viene prodotto.
>
> [p. 10]

Contrariamente a quanto ipotizzato in un primo approccio alla lingua come universo statistico, soprattutto in Giraud (1954), approccio per il quale le unità del sistema linguistico sono caratterizzate oltre che dalle opposizioni e relazioni «qualitative» che determinano la struttura del sistema, anche dalle loro rispettive probabilità di uso testimoniate dalle occorrenze delle varie unità nei testi (parlati e scritti), con frequenze relativamente stabili, gli stessi dati ottenuti dal LIF facevano registrare l'esistenza di variazioni

nella frequenza di determinate classi di parole nei vari sottoinsiemi esaminati: variazioni strettamente correlate alla natura di tali sottoinsiemi. Ciò confermava l'ipotesi già presente nel *Français Fondamental* secondo cui le frequenze delle parole in genere non sono stabili, ma variano in relazione allo «stile» ed al «tema». La concezione generale cui si ispirano considerazioni di questo tipo è quella per cui, come ad esempio in Moreau (1962), il concetto di probabilità, come carattere intrinseco dell'unità del sistema linguistico, non è espresso: l'universo statistico è definito piuttosto come l'insieme di tutti i testi (parlati e scritti) prodotti in un certo periodo di tempo. Gli autori del LIF chiudevano il loro contributo affermando che:

> L'effettivo riconoscimento degli strati di lingua dovrà derivare da un lavoro accurato e documentato di induzione da spogli sempre più numerosi […] In questa stessa direzione gli spogli e le elaborazioni per il nostro lessico di frequenza dell'italiano contemporaneo costituiscono un primo contributo allo studio statistico della lingua italiana come universo lessicale.
>
> [p. 15]

Immediatamente successiva al LIF, la pubblicazione di Juilland-Traversa (1973). Anche in questo caso il *corpus* è rappresentato da circa 500000 parole localizzate lungo cinque sottoinsiemi di circa 100000 parole ognuno, appartenenti a cinque generi di scritto differenti: teatro, romanzi, novelle e storie brevi, saggistica, periodici e quotidiani, letteratura tecnico-specialistica, per un arco di tempo che va dal 1920 e al 1944. È evidente come, rispetto al LIF, esso sia quasi altrettanto sincronico, ma meno attuale. Ancora dedicato all'italiano scritto il VELI 1989 che presenta un *corpus* con un'estensione vastissima, circa 26 milioni di occorrenze, ma con un orientamento molto forte verso lo scritto formale, data la tipologia testuale prescelta, costituita fondamentalmente da settimanali informativi, settimanali economici, notiziari ANSA e simili. Unico caso di analisi di frequenza dedicata al parlato, quello del KBS (1991), contraddistinto, però, e dalla relativa limitatezza dell'insieme delle occorrenze, e dagli obiettivi squisitamente didattici che hanno spinto a definire una scelta di testi non tanto su una possibile tipologia del parlato, quanto su una serie di criteri legati all'esperienza dell'insegnamento dell'italiano come lingua straniera.

1.2. Il Lessico di Frequenza dell'Italiano Parlato

1.2.1. Finalità del LIP

De Mauro, nel saggio introduttivo («Gli obiettivi della ricerca») al LIP chiarisce immediatamente che:

> Il primo obiettivo della ricerca è accertare o, meglio, predisporre preliminari indispensabili strumenti e documenti obiettivi per accertare in che

> misura l'uso parlato dell'italiano presenti tali caratteristiche, e in che misura resti invece tuttora ancorato (rispetto agli usi parlati di altre lingue) a caratteristiche dell'uso scritto ovvero, alternativamente, a caratteristiche di parlato bensì, ma di parlato orientato sul dialetto locale. In altri termini, l'obiettivo è mettere a fuoco l'esistenza di un nucleo di convergenza standard degli usi parlati colti nella loro variabilità regionale e di genere.
>
> [p. 25-26]

In effetti, il LIP costituisce il primo lessico di frequenza dell'italiano parlato di dimensioni «standard», data la taglia del campione preso in esame, e rappresenta, in linea di massima, uno strumento volto alla verifica sperimentale delle ipotesi relative ai processi di unificazione linguistica dell'Italia, con particolare attenzione, alla verifica della diffusione dell'italiano nel suo registro parlato:

> Da qualche anno per la prima volta dopo secoli ci troviamo dinanzi a un uso parlato dell'italiano sufficientemente abituale e largo, cioè non limitato regionalmente a Toscana e Roma, né limitato per generi di discorso e contenuti soltanto alle occasioni più formali. Di qui l'ipotesi di potere cogliere anche nel parlato italiano quelle caratteristiche che, dopo gli studi di Charles Bally e Henry Frei sulle caratteristiche specifiche e differenziali degli usi parlati rispetto agli scritti e dopo le ampie ricerche sul «francese fondamentale», dobbiamo attenderci come tipiche di un uso «parlato-parlato». Esse sono, per citarne solo alcune: il privilegio accordato alle parole di più alto uso (contro la maggiore varietà lessicale dello scritto, che premia le basse frequenze); la presenza più accentuata di colloquialismi (si intende: di colloquialismi non dialettali o vernacolari, ma comuni e standard); una semplificazione dell'apparato morfo-sintattico del verbo; il privilegio accordato ai nessi sintattici di più alto uso.
>
> [p. 25]

A quest'ipotesi di fondo si collegano due sub-ipotesi e cioè che certamente gli usi parlati dell'italiano si differenzino, da un lato, sull'asse del maggiore o minore grado di improvvisazione o progettazione dei discorsi e di condizioni più o meno favorevoli all'informalità (intesa come appello alle circostanze non linguistiche dello scambio comunicativo) e, dall'altro, sull'asse del grado maggiore o minore di distacco dalle caratteristiche puramente locali, municipali, del parlato, che, in genere, dovrebbero essere influenzate da tratti dialettali e regionali. L'italiano, insomma, oggi divenuto una lingua come altre lingue storico-naturali, in un processo che, per la prima volta nella storia moderna della penisola, vede realizzata una sostanziale unità, sia, naturalmente sul versante delle realizzazioni scritte, sia su quello delle realizzazioni parlate. E ciò grazie a quei processi esterni (grandi migrazioni interne, crescita dei livelli di istruzione e scolarizzazione, diffusione delle comunicazioni di massa, in particolare radio-televisive) che venivano individuati già in De Mauro (1963).

1.2.2. Il *corpus* del LIP

Il LIP è costituito da un *corpus* di circa 500000 parole grafiche ed è stato detto, per tale ragione, di «dimensioni standard», essendo questa la taglia del campione ritenuta ottimale. Il numero delle occorrenze di forme grafiche è di 475883, mentre il numero di occorrenze di lemmi è di 496335. Questa differenza è dovuta al particolare trattamento riservato alle preposizioni articolate e ai clitici. Il dato più interessante, però, è da ritenersi quello relativo alla rappresentatività geolinguistica e testuale del *corpus*.

Dal punto di vista della rappresentatività geolinguistica, la costituzione del *corpus* del LIP ha posto in partenza numerosi problemi. Infatti, data l'impossibilità della riduzione del prelievo del campione in un'unica area — per le ragioni che la storia linguistica italiana ha tradizionalmente posto e che sono ben descritte nella letteratura specialistica di tipo storico-linguistico e sociolinguistico — le scelte in alternativa fra loro erano in partenza diverse. Una raccolta svolta in tutti i grandi centri urbani avrebbe, però, presentato comunque lo svantaggio di essere sì qualitativamente rappresentativa, ma con dati statisticamente poco significativi, dato che almeno una dozzina sarebbero stati i centri urbani da prendere in considerazione. D'altro lato, la limitazione di tale raccolta ai centri più importanti dal punto di vista demografico (Roma, Milano, Napoli e poi Torino, Genova e Palermo) avrebbe portato ad escludere un centro come Firenze che, indipendentemente dal ruolo svolto nella storia linguistica italiana, ancora oggi presenta la più alta concentrazione di italofoni. D'altronde:

> Il riferimento ad una valutazione siffatta, se porta ad includere in primo luogo Firenze, certamente porta ad includere subito dopo anche Roma […] quindi Milano, per la nota apprezzabilità del suo standard, infine Napoli, per l'antica accettazione che l'italofonia vi ha avuto nei ceti colti e borghesi e per il grande prestigio che espressioni inizialmente napoletane hanno avuto nel costituirsi delle fasce affettive, basse, se si vuole volgari, del nostro vocabolario comune.
>
> [p. 31]

Infine, l'aggiunta di un quinto centro, pur statisticamente possibile, avrebbe, nell'opinione degli autori del LIP, sbilanciato la rappresentatività media complessiva del campione, a vantaggio o del Nord (con l'aggiunta di Torino) o del Sud (con l'aggiunta, per esempio, di Palermo). La riduzione, quindi, dei centri urbani a quattro — Milano, Firenze, Roma e Napoli — che vede confluire il criterio demografico e quello del prestigio linguistico, permette di ottenere una raccolta di materiali equilibrata per ciò che concerne la documentazione centrale toscana e non toscana, romana, settentrionale e meridionale.

Relativamente più semplice la questione della rappresentatività testuale, cioè quella riguardante i generi del parlato. Partendo da una definizione di «parlato» come produzione fonica, spontanea, cioè non basata sull'esecuzione di un testo scritto precostituito, il LIP, secondo M. Voghera («Le scelte per la

costituzione del corpus») non raccoglie una varietà specifica dell'italiano, ma riunisce testi di italiano «tendenzialmente comune e unitario parlato in tutto il territorio nazionale». Sono stati, quindi, individuati cinque diversi gradi di una polarità comunicativa procedente da una maggiore ad una minore naturalezza e, per ognuno di esso, sono state spogliate 100000 parole, 25000 per ogni città del campione:

1. Scambio comunicativo bidirezionale con presa di parola libera faccia a faccia in: conversazioni in casa; conversazioni sul luogo di lavoro (tra pari grado, tra non pari grado, con estranei in uffici); conversazioni nell'ambito scolastico e universitario; conversazioni in luoghi ricreativi e sui mezzi di trasporto.
2. Scambio comunicativo bidirezionale con presa di parola libera non faccia a faccia in: conversazioni telefoniche normali; conversazioni telefoniche registrate alla radio; messaggi registrati nelle segreterie telefoniche.
3. Scambio comunicativo bidirezionale con presa di parola non libera faccia a faccia in: assemblee legislative; dibattiti culturali; assemblee studentesche; assemblee sindacali; incontri di lavoratori; interrogazioni nella scuola elementare; interrogazioni nella scuola secondaria; esami universitari; interrogatori processuali; interviste alla radio e alla televisione.
4. Scambio comunicativo unidirezionale in presenza del/i destinatario/i in: lezioni di scuola elementare; lezioni di scuola secondaria; lezioni universitarie; relazioni a congressi o convegni politici e sindacali; relazioni a congressi o convegni scientifici; comizi politici; omelie; conferenze non specialistiche; arringhe giudiziarie.
5. Scambio comunicativo unidirezionale a distanza o differito su testo non scritto in: trasmissioni televisive; trasmissioni radiofoniche.

Le linee di ricerca che una rappresentatività testuale di tale tipo offre sono in partenza molteplici, ma, in primo luogo, essa offre la possibilità di sviluppare un confronto tra la distribuzione delle frequenze calcolate sull'intero *corpus* e la distribuzione delle frequenze relative ad ogni genere o tipo testuale/comunicativo particolare.

1.3. Il lessico dell'italiano parlato

Il confronto, in particolare, tra il LIP ed il LIF[4] confermerebbe, nell'opinione di M. Vedovelli («Confronti tra il LIP e le altre liste di frequenza dell'italiano»), da un lato, la «povertà lessicale del parlato», ipotizzata in più studi e per l'italiano in De Mauro (1970), comprovata dal fatto che la testa di lista del

4. Questo genere di confronti è comunque problematico, non tanto per gli apparati statistici, quanto almeno per le parzialmente diverse procedure di lemmatizzazione adottate. Ad esempio, il LIF impiegava criteri di lemmatizzazione grammaticale tale che spesso unificava sotto la stessa unità lessicale elementi, invece, differenziabili per categoria grammaticale, causando una non semplice distinguibilità dei dati per frequenza, valore d'uso e rango.

LIP copre con un numero inferiore di lemmi un'area di testi più vasta — in pratica il parlato utilizza un minor numero di parole diverse — e dal fatto che non si registrano ampie, sistematiche e specifiche nuove entrate nella «testa» del LIP rispetto al LIF; e, dall'altro, conferma che, aldilà di un'iniziale più ampia capacità di copertura del LIP, le liste elaborate mostrano poi un bilancio per molti versi in pareggio. Anche se comunque ci sono delle differenze. Lo spostamento verso l'alto (rispetto al LIF) riguarda in modo particolare i verbi, specialmente gli «operatori di frase», cioè verbi che, nella terminologia utilizzata dagli autori, sembrano perdere specificità semantica per caricarsi anche di una funzione sintattico-pragmatica nel testo parlato. Alle variazioni registrate verso l'alto, come ad esempio:

	LIP	**LIF**
dire	16	23
andare	26	39
dovere	36	38
stare	48	50
parlare	62	72

Si contrappongono, però, anche talune modificazioni del rango d'uso verso il basso, come ad esempio:

	LIP	**LIF**
trovare	83	68
prendere	96	76
credere	106	103
ricordare	116	107
chiedere	122	117

Altresì si assiste ad un innalzamento sistematico, per esempio, del rango degli avverbi (ci sembra soprattutto di modo): esso può essere interpretato in direzione della specificità del parlato rispetto allo scritto, nel senso, però, di un impoverimento semantico degli elementi lessicali del parlato. Questi vengono a perdere parzialmente il loro valore semantico «pieno» per assumere una funzione che contribuisce alla strutturazione specifica del testo parlato. Dal confronto tra LIP, LIF e lo stesso VELI emerge che:

> A parità sostanziale dei contenuti lessicali delle teste delle liste […] ciò che sembra caratterizzare autonomamente il LIP è la ristrutturazione delle posizioni del rango d'uso, con la tendenza nello spostamento verso le posizioni superiori rispetto al VELI. Rispetto al LIF, invece, si hanno slittamenti sbilanciati verso l'alto e verso il basso: sulle prime 450 parole della lista per rango d'uso del LIP, il 55% si sposta verso le posizioni superiori, mentre il 43% si sposta verso il basso; l'1,5% rimane nello stesso rango (e sono lemmi che rientrano nei primissimi).
>
> [p. 128]

Nel caso di sostantivi e aggettivi la variazione tende a divenire più forte e la spinta verso la testa della frequenza si fa più debole. Valutando complessivamente i dati emersi dallo spoglio, Vedovelli nota come il confronto fra le liste non metta in luce nuovi inserimenti del parlato rispetto allo scritto, e come non sia possibile individuare aree semantiche privilegiate in confronto allo scritto. Colpisce la conferma della difficoltà di isolare un'area lessicale specifica ed esclusiva del parlato: tra le nuove entrate si possono rilevare solo le formule standardizzate di attacco, conferma e fine del discorso:

> Poi si registra da un lato un'emersione di lemmi e dall'altro un abbassamento delle posizioni. Come interpretare tale contraddittoria dinamica? In De Mauro et alii (1993a) avevamo tentato una interpretazione che collocava il fenomeno della emersione di lemmi come specificità del parlato in relazione da un lato alle strategie psicolinguistiche di gestione dell'interazione orale, dall'altro in rapporto ai caratteri semantici degli elementi spostati verso l'alto, e infine ricollegandoci all'analisi di Simone (1990) dei meccanismi testuali specifici del parlato (l'effetto copia e l'effetto quasi-copia). […] Ora, con i dati globali del LIP e con diverse altre fonti sulle quali è stato operato il confronto, quel tipo di interpretazione non viene comunque sconfessata, ma sicuramente limitata alle emersioni, le quali peraltro rappresentano solo una tendenza e non quella predominante nettamente. […] Sul piano funzionale l'innalzamento di posizione di questi lemmi (e in modo particolare dei verbi operatori di frase, o più propriamente di «testo parlato») deriva da una loro parziale desemantizzazione specifica e dalla loro trasposizione in elementi con valore di snodo nella struttura del testo parlato. Se si considerano gli elementi «pieni», l'innalzamento ha per conseguenza l'estensione di area semantica a scapito della specificazione, e tale fenomeno ha un valore funzionale in rapporto alla dimensione psicolinguistica dell'interazione orale. Nel parlato, infatti, la possibilità di controllo metalinguistico e di selezione degli elementi lessicali «pieni» si attua entro condizioni «ristrette» che devono consentire l'ottimizzazione della fissazione e trasmissione del senso in rapporto alle modalità dello scambio in battute o della progettazione rapida passo a passo del discorso. Le condizioni di interazione, pertanto, spingono a selezionare gli elementi con più immediata e vasta applicabilità semantica, e questi vengono ad avere posizioni più alte nella lista di frequenza/uso.
>
> [p. 135]

A quali conclusioni generali e storico-linguistiche si può giungere a partire dai dati emersi dal LIP? De Mauro, ricordando le ipotesi di partenza e cioè che la diversità tra scritto e parlato vive non solo a livello di «materia», ma anche a livello di «uso» e di «norma» e che i caratteri della storia linguistica italiana per i quali, al di fuori di Roma e dell'area fiorentina, la lingua italiana è stata parlata così variamente per generi di discorso ed estesamente solo da breve tempo, mette in evidenza come entrambe siano state verificate. In realtà, l'uso parlato, potendo avvalersi del supporto della prosodia, dell'immediata correlazione con ciò che è extra-linguistico, della possibi-

lità di immediata corregibilità di quanto viene detto, segue «scorciatoie» che, invece, l'uso scritto non può intraprendere. Tutto ciò si associa a realtà lessico-statistiche significative e cioè:

a) il parlato può essere più ricco di forme morfosintatticamente ambigue, cioè interpretabili come appartenenti a più di una classe di parole;
b) parlato e scritto ricorrono in modo differenziato nelle varie categorie morfosintattiche: nel secondo prevalgono sostantivi e aggettivi, che insieme costituiscono il 38,7%, seguiti percentualmente da preposizioni, verbi, etc.; nel primo, invece, dominano i verbi e i pronomi, gli operatori di apertura, conferma e controllo del discorso cioè avverbi e congiunzioni, mentre sostantivi e aggettivi sono più rari, costituendo insieme il 24,5%.

Il dato comunque più rilevante è costituito dal fatto che il parlato utilizza maggiormente un lessico meno ricco di lemmi differenti, mentre lo scritto per raggiungere una «pari informatività», cioè una pari copertura, deve attingere ad un patrimonio più vasto di forme singolarmente e mediamente di frequenza inferiore. L'ultima più importante considerazione concerne le condizioni storico-linguistiche: il LIP fornisce i dati sperimentali per misurare e valutare «il processo collettivo di convergenza» verso l'italiano e verso l'unificazione e l'omogeneizzazione linguistica del paese:

> L'italiano non serve più solo per trattati filosofici o per cantare romanze, ma fascia e innerva l'esperienza quotidiana, volta a volta umile e alta, intima e privata, pubblica e collettiva, scientifica e letteraria, di decine e decine di milioni di persone.
>
> [p. 160]

1.4. L'analisi lessico-grammaticale del corpus del LIP

Un *corpus* come quello del LIP che, oltretutto, consente un accesso diretto ai testi che ne costituiscono la base documentaria, si offre come strumento fondamentale per la descrizione dell'italiano d'uso. In D'Agostino (1992) abbiamo voluto intraprendere una prospettiva di ricerca che proiettasse la descrizione lessico-grammaticale sull'analisi dei testi, in un approccio che si è definito di «sociolinguistica interna». Successivamente, in D'Agostino (1993), è stata nostra intenzione battere la strada della verifica sperimentale sul piano dell'analisi delle fonti dell'uso parlato. Il nostro punto di partenza è stato, comunque, sempre quello per il quale, distinta la lista delle unità lessicali di una lingua individuate sul piano morfo-fonologico da quella individuata sul piano sintattico-semantico, costituendo quest'ultima un insieme di ben più vaste dimensioni di quello rappresentato dalla prima, le singole entrate lessicali sono analizzabili e classificabili, sul piano dell'uso sintattico-semantico, in base ai principi della selezione e della co-occorrenza. Si è rivelato di interesse nella nostra prospettiva l'ana-

lisi della frequenza di distribuzione e delle variazioni di questa dei vari operatori, sia di tipo verbale sia di tipo non verbale, come delle costruzioni a supporto, cioè la distribuzione delle varie forme di frase e delle proprietà ad esse associate. Dall'esame comparativo dei risultati delle analisi di frequenza delle fonti dello scritto e del parlato, colpisce particolarmente la diversa distribuzione delle categorie morfosintattiche negli usi scritti in confronto a quelli parlati, come ha sottolineato De Mauro nelle conclusioni al LIP. Infatti, registriamo un quadro percentuale di tale tipo:

uso	**parlato**	**uso**	**scritto**
V	20,0%	*N*	21,7%
N	15,7%	*Prep*	17,2%
Prep	11,6%	*Agg*	17,0%
Pro	10,9%	*V*	10,4%
Cong	10,1%	*Cong*	4,3%
Avv	10,1%	*Avv*	3,8%
Agg	8,8%	*Pro*	2,5%

Ma i dati, naturalmente, concernono unicamente la lista dei lemmi e non illustrano la funzione sintattico-semantica che essi assumono, nella predicazione, nei vari contesti frastici.

Se si esclude il peso quantitativo delle occorrenze dovute al meccanismo della «ripetizione dialogica», che con varie tipologie contraddistingue in quanto tale l'interazione parlata e, se non può stupire la variazione percentuale di membri di categorie come *Pro*[5] e *Avv*,[6] è certamente significativo che ci sia una variazione percentuale della frequenza dei verbi tra scritto e parlato, o di quella dei nomi e degli aggettivi, ma c'è da chiedersi altresì quale possa essere il «peso» qualitativo di tali differenze, non più dal punto di vista soltanto della lista dei lemmi, ma da quello della funzione da essi svolta nei contesti frastici di occorrenza. Siamo convinti, infatti, che un riesame dei dati statistici forniti dal LIP alla luce di una considerazione fun-

5. L'innalzamento in assoluto dei *Pro* non è dato significativo in se stesso. È necessario, infatti, ricordare come, all'interno del più generale meccanismo della deissi, alcune forme pronominali effettuino un balzo in avanti nell'uso parlato: *io* e *tu* coincidenti con le funzioni di soggetto dell'enunciazione e di ascoltatore, e tutte quelle altre forme esoforiche (*questo*, *quello,* ecc.) che, per definizione, sono proprie dell'atto comunicativo nella sua realizzazione parlata.
6. Per giustificare l'innalzamento della quantità di forme avverbiali, bisogna ricorrere nuovamente alla considerazione della co-occorrenza. Infatti, una vasta tipologia di avverbi, in contesti frastici propri del parlato, finisce con il perdere lo statuto di *Avv* e assume su di sé il valore di puro e semplice segnale discorsivo: *cioè*, *praticamente*, *insomma* ecc. In tal senso, una volta ancora, il calcolo basato sulla pura considerazione per lemmi bruti risulta falsato e ingannevole. Naturalmente, ciò vale anche per membri di altre classi di parole, ad esempio i verbi in sequenze come: *dai, vieni pure tu*!, *mi consenta…*, *veda bene…*, *faccia attenzione…*

zionale delle varie occorrenze nei diversi contesti possa contribuire ad ancora meglio definire i caratteri fondamentali dell'uso parlato dell'italiano contemporaneo, magari approfondendo, o anche in parte correggendo, le ipotesi e le stesse conclusioni cui gli autori del LIP sono pervenuti. Infatti, è nostra opinione che la questione fondamentale, dal punto di vista della definizione degli usi lessicali propri del registro parlato (come del resto di quello scritto), stia nell'individuazione, non solo di una lista di lemmi, ma anche delle occorrenze di tali lemmi all'interno di forme di frase che ne evidenzino le proprietà definizionali sul piano del Senso come della Forma. D'altronde, soltanto un'indagine del genere può confermare su vasta scala documentaria le stesse affermazioni della sociolinguistica italiana sulla definizione di un nuovo standard e sui caratteri, in particolare morfo-sintattici, che questo viene ad assumere. Ed in questa direzione la documentazione offerta dal LIP ci pare costituire una fonte decisiva.

1.4.1. L'analisi di *fare* e *dare*

La tesi principale già sostenuta in precedenza[7] a proposito degli esempi forniti dai lemmi *fare* e *dare* nel LIP può essere così riassunta: un'analisi lessico-grammaticale mostra le variazioni diatopiche e diastratiche meglio di una considerazione per lemmi bruti. L'esame di tali variazioni diviene più significativo, quando sono esplicitati gli usi sintattico-semantici delle varie forme. Del resto, una considerazione degli usi si mostra pertinente ogni qual volta si prende la strada dell'analisi del vocabolario di una lingua. Richiamando la più volte sottolineata distinzione tra lista delle forme e lista degli usi, e sapendo che, per una lingua come l'italiano, quest'ultima è mediamente più grande della prima da cinque a dieci volte, ci pare del tutto evidente come, focalizzata l'attenzione sull'analisi lessicale dell'uso linguistico, si giunga sempre poi alla necessità di indagare nella direzione della selezione e della co-occorrenza. Essere in possesso dei dati sul fatto che un lemma sia noto e compreso, ad esempio, oppure sul fatto che abbia una certa frequenza è fuor di ogni dubbio rilevante, ma crediamo che lo sia maggiormente avere delle conoscenze sperimentali sull'uso che di questo il parlante fa. Ciò che egli fa è utilizzare non lemmi isolati, ma cogliere insieme aspetti sintattici ed aspetti semantici, occorrenze in contesti frastici, che, in particolare, nel parlato sono contraddistinti da cancellazioni, variazioni d'ordine, riprese, cambi di strategia e di argomento, etc. Tutti fenomeni legati per definizione al «formato» proprio del parlato. In realtà, dal nostro punto di vista, nel parlato come nello scritto l'unità minima del lessico non è una singola forma, ma comunque una frase o, meglio ancora, un gruppo coerente di frasi accomunate dalla relazione esistente tra un determinato operatore ed i suoi argomenti che, indipendentemente dalle realizzazioni concrete, rimane la stessa. Relazione, questa, che va ad inserirsi in reti funzionali che abbiamo altrove definito «costellazioni parafrastiche». Ed è in tal

7. Cfr. D'Agostino (1993).

senso che, in questa sede, abbiamo parlato del riconoscimento delle costruzioni, delle forme di frase possibili, di una loro tipologia associata a dei lemmi e della loro distribuzione lungo frequenze d'uso diverse e lungo tipi o generi discorsivi differenti.

Alcune considerazioni generali vanno fatte a partire dai due lemmi analizzati. In primo luogo, va sottolineato ancora una volta nettamente che in base ai dati concernenti l'uso da parte dei parlanti di *fare* e di *dare* messi a confronto, si registra il fatto che il secondo è, in misura più netta del primo, elemento lessicale «pieno». Una volta individuata una tipologia strutturale di *dare*, cioè le costruzioni alternative possibili messe a disposizione dal sistema dell'italiano contemporaneo, si nota come i parlanti impieghino più frequentemente tale verbo come operatore più che come strumento funzionale, in grado maggiore di quanto facciano con *fare*. Questa considerazione ci appare importante poiché è una riprova del fatto che una considerazione in lemmi bruti ci illumina meno sia sul piano della definizione di un sistema linguistico, sia su quello dell'uso che i parlanti ne fanno.

In secondo luogo, però, sia nel caso di *fare* che in quello di *dare*, nei vari contesti frastici in cui essi occorrono, assistiamo al fatto che essi sono nell'assoluta maggioranza dei casi — più *fare* e meno *dare*, naturalmente — elementi fondamentalmente funzionali, poiché dire «più o meno vuoti» sarebbe semplicistico, dato il peso qualitativamente assai significativo delle costruzioni a supporto e delle costruzioni causative.[8] Tradotta in termini di uso linguistico, tale affermazione significa che i parlanti utilizzano i due verbi fondamentalmente come elementi funzionali, più che come elementi lessicali «pieni», (ossia come operatori), anche se va notato che, dal confronto fra i due, emerge che *fare* è in primo luogo utilizzato come verbo supporto e che le «maggioranze» che essi insieme costituiscono sono in un caso «assoluta» e nell'altro «relativa». Le costruzioni causative, per utilizzare un gergo sportivo, occupano soltanto la terza posizione.

Si potrebbe, quindi, ipotizzare, anche se ciò richiede un approfondimento di analisi, che negli usi del parlato si assista ad una riorganizzazione della predicazione intorno a nuclei nominali di tipi di frase abitualmente accomunati intorno ai verbi supporto.

2. Il LIP come strumento

Date queste premesse, il LIP, in ogni caso, si rivela in principio uno strumento fondamentale:

— per la definizione generale del «formato» parlato di una lingua; infatti, ad esso soggiace un'ipotesi forte e cioè che il parlato si differenzi da altre

8. Per quanto riguarda le nozioni di «verbo supporto» e di «costruzione causativa» si veda, almeno, D'Agostino (1992) e (1993).

modalità di costruzione e trasmissione dell'informazione, più che per certi tratti di natura sintattico-testuale, soprattutto perché trae la sua peculiarità da condizioni semiotiche inerenti al rapporto emittente-destinatario; il vero discrimine tra i vari tipi di parlato è la presenza/assenza di turni di parola;
— per l'analisi sociolinguistica: infatti, esso costituisce uno strumento di verifica delle ipotesi più volte avanzate in letteratura sulla configurazione di un «neo-standard», in quanto la definizione stessa dei tratti che caratterizzerebbero l'evoluzione dell'italiano contemporaneo necessita di una prova documentaria non casuale o forzata, ma quantitativamente e qualitativamente fondata;[9]
— infine, per qualsiasi analisi documentaria sull'italiano fondamentale, a integrazione del LIF, e, quindi, per tutte le applicazioni didattiche.

3. Il LIP come strumento didattico

Ed è per l'appunto nella dimensione dell'utilizzabilità del LIP dal punto di vista didattico che riprenderemo, qui di seguito, alcuni aspetti tra i vari possibili.

3.1. La varietà della gamma del parlato

In primo luogo, si possono ricordare, ad esempio, le posizioni come quella di Gensini (1994) che sottolinea l'importanza del LIP per una attenzione «nuova» al parlato in relazione alle varie situazioni comunicative. Gensini, riferendosi ai materiali che costituiscono il corpus del LIP, ricorda come, tali materiali, procedendo lungo una scala di decrescente «naturalezza», vadano:

> Dalle comunicazioni di massima libertà espressiva e comunicativa, la conversazione faccia a faccia senza turni di parola prestabiliti, a situazioni sempre più vincolate, sia di tipo dialogico, sia di tipo monologico.
>
> [p. 195]

Consentendo, con funzione di base documentaria, di modificare il piatto atteggiamento scolastico tradizionale nei confronti del parlato generatore di una situazione dalla quale:

9. Un esempio valga per tutti: il cosiddetto «*che* polivalente» citato in tutta la manualistica sociolinguistica italiana come connettivo generico di forte diffusione. Ad un primo esame del campione romano del LIP — cioè del campione che con quello napoletano presenta il più alto tasso di dialettalismi dell'intero corpus — come esempio principe dell'influenza di caratteri del parlato sul neo-standard quantitativamente appare, se non marginale, almeno fortemente minoritario rispetto agli altri impieghi possibili del *che*.

> Discende anche una significativa perdita di agilità comunicativa, dirò così, nel passare da un tipo all'altro di situazione di uso orale della lingua, quali ad esempio il parlare su base di appunti, il parlare riferendosi a un testo scritto, il parlare a braccio su un tema dato. Lo «stile» di parlato ritrovabile in molti studenti mostra un carattere indifferenziato, o per meglio dire, una sorta di polarizzazione acritica tra quelle forme di parlato-parlato che sono proprie della comunicazione quotidiana più informale e un registro paludato, quasi di recitazione del discorso, che tende a sottrarsi al dialogo e alla verifica.
>
> [p. 197]

Dunque, in realtà, il corpus LIP si pone come raccolta di materiale autentico di vario tipo funzionale utilizzabile in una pratica didattica attenta alla dimensione e della comunicazione reale e di quella propriamente orale.

3.2. Vocabolario di base per stranieri

In secondo luogo, la più immediata applicazione si orienta verso la didattica dell'italiano lingua straniera. Come sottolinea Stammerjohann (1994):

> Per noi stranieri, insegnanti della lingua italiana, ovvero per noi insegnanti dell'italiano come lingua straniera, il LIP rappresenta la più potente base di riferimento per determinare il lessico da insegnare, specie il lessico da insegnare prima, cioè il lessico, o vocabolario fondamentale.
>
> [p. 83]

In particolare allo scopo di:

a) decidere quali liste di frequenza ora esistenti rispondono più ai bisogni dell'insegnamento della lingua come lingua straniera;
b) collazionare le liste (selezione di 2500 lemmi);
c) collazionare i vocabolari di base;
d) collazionare i lessici dei corsi di lingua italiana più usati;
e) fare un'analisi differenziata del livello soglia, quantificando in modo particolare il numero di atti linguistici in cui una parola occorre.

3.3. Quale standard insegnare?

In terzo luogo, si apre la questione più spinosa sia dal punto di vista teorico che da quello applicativo: lo standard di lingua preso a modello e utilizzato nella costruzione dei materiali didattici. Senza voler entrare nel merito del dibattito sulle nozioni di «standard» e di «neo-standard», per le quali si rimanda alla bibliografia di sociolinguistica citata, ci basti qui evidenziare come strumenti del tipo di quello cui ci si riferisce costituiscano la fonte documentaria per la scelta delle strutture linguistiche di base — si faccia attenzione: «di base» soltanto! — sulle quali sviluppare la compo-

nente più formale della didattica dell'italiano. A tale proposito si possono, naturalmente, fornire esempi diversi. In questa sede, la scelta è certamente parziale, ma non casuale perché fondata su dati empirici ampiamente attestati.

3.3.1. Pronomi personali

Ad esempio, seguendo Renzi (1994), l'attenzione può essere rivolta alla sorte che è capitata ai paradigmi pronominali: le forme personali. Sotto il lemma *egli* con 39 occorrenze appare anche la forma *lui* che realizza ben 764 occorrenze. Inoltre, ben 33 occorrenze del primo sono localizzate nel contesto (D) del LIP, cioè quello più formale. Il primo è, dunque, praticamente morto. Ancora in maniera più evidente la stessa sorte è toccata alla forma *ella* che non registra alcuna occorrenza. Le indicazioni didattiche sono del tutto evidenti.

3.3.2. L'uso dei tempi verbali

Analoga attenzione, può essere rivolta all'uso, ad esempio, dei tempi verbali. In tal senso, in Gambarara (1994) si registra la prova del disuso del passato remoto. Si possono, inoltre, verificare empiricamente le analisi fatte da altri, sulla base di altri corpora, sui tempi del parlato italiano: ad esempio quanto sostenuto da Bazzanella (1994)[10] sul rapporto futuro/presente e sugli usi dell'imperfetto. In questo caso, spostandoci sul complesso asse dell'interazione tempo/modo/aspetto, possono essere analizzati e verificati esempi come quelli sinteticamente presentati nei paragrafi seguenti, che rappresentano altrettanti casi di problemi legati all'italiano lingua straniera.

3.3.2.1. Il futuro

Rimandando a Bertinetto (1986) e (1991), per quanto concerne la definizione generale della problematica inerente ai tempi verbali dell'italiano, si può procedere ad un esame dei fenomeni di modificazione dell'uso dei tempi e degli impieghi non deittici di questi:

Futuro con Presente:
A. *Vieni più tardi?*
B. *Sì, passo alle cinque*

Futuro con Forme perifrastiche:
stare per Vinf =:
senti, in che cosa ti stai per maturare
andare a Vinf =:
dopo di che andiamo ad accontentare la nostra ascoltatrice

10. Il cui corpus è limitato e raccolto «artigianalmente» e non si presta ad analisi automatiche.

Uso epistemico:

A. *che ore sono?*
B. *più o meno saranno le otto*

e la relazione con parametri come:

Commitment:
adesso cercherò di essere più chiaro

Priming:
chissà che cosa costruirai, che cosa disegnerai e che cosa poi verrà costruito

o il futuro nelle subordinazioni:

dipende se lo farai
la avverto che se avrà conseguenze lei sarà la responsabile

o i cosiddetti usi modali:

si scriverà o si telefonerà solo in alcuni momenti
firmerà questa petizione
l'acqua bollirà a 100 gradi

3.3.2.2. L'imperfetto

Analogamente, per quanto concerne l'imperfetto, tempo, definito normalmente «buono per tutti gli usi», nel senso che vede realizzate un altissimo numero di occorrenze con valori sia di tipo deittico sia non deittico, è possibile andare a verificare i concreti impieghi sperimentalmente documentati e la loro diffusione:

Imperfetto ipotetico:

se potevo, venivo

Imperfetto ludico:

facciamo che io ero il lupo e tu eri Cappuccetto Rosso

Imperfetto di cortesia:

A. *Desiderava?*
B. *Volevo un chilo di pane*
guardavo solo se c'era un posto

Imperfetto epistemico-doxastico:

cosa c'era domani al cinema?

Imperfetto di pianificazione:

A. *Non puoi farlo domani?*
B. *Domani andavo in biblioteca*

3.3.3. Fenomeni sintattici

Più volte sono stati segnalati anche come tratti propri della nuova configurazione dello standard scritto fenomeni in origine propri del parlato. Ad esempio, in Sobrero (1993) e negli altri titoli qui citati ci si sofferma su fenomeni come quelli riportati qui dappresso.

3.3.3.1. Dislocazioni

Il problema delle dislocazioni, delle frasi scisse e pseudo-scisse, appartenendo queste al dominio della «marcatezza», si intreccia con le questioni tema/rema e dato/nuovo. Si possono ricordare gli esempi in Bazzanella (1994) come:

la mia dichiarazione dei redditi non l'abbiamo potuta fare
senti, c'è una materia che proprio detesti
sono medici, insegnanti, professionisti che hanno sulle spalle un morto

3.3.3.2. Il passivo

Lo stesso meccanismo del passivo — dominio anch'esso proprio della marcatezza alla stessa stregua delle dislocazioni e della stessa diatesi neutra — può essere sottoposto a verifica, soprattutto una volta che si sia messa da parte l'ipotesi strettamente diamesica che parte dalla presunzione di forte correlazione tra uso scritto e passivo. Un controesempio chiaro è quello citato, ad esempio, in Bazzanella (1994) delle radiocronache e telecronache sportive:

è Bagni che ora viene attaccato da Galia[11]
il pallone adesso viene indirizzato verso Maradona

In particolare, poi, gli usi del passivo si ricollegano a due altri principali questioni: la correlazione con il tipo di predicato che attiva la costruzione passiva da un lato, e, dall'altro, quella delle possibili cancellazioni della posizione «agente».

4. Conclusioni

In questa sede — naturalmente in maniera molto sintetica — abbiamo cercato di illustrare come uno strumento nato a scopi descrittivi della situazione attuale dell'italiano parlato si presti allo stesso tempo sia a considerazioni di ordine generale rivolte alla definizione stessa del formato «parlato» della lingua, sia a riflessioni applicative in particolare di tipo didattico, costituendo, a nostro avviso, la base documentaria necessaria per la pratica di insegnamento, nei suoi vari aspetti.

11. In questo esempio alla forma marcata passiva si aggiunge l'altro meccanismo della frase scissa.

Bibliografia

AA.VV. (1993). *Atti del XXV Congresso Internazional della S.L.I.* Roma: Bulzoni.

BAZZANELLA, C. (1994). *Le facce del parlare*. Firenze: Nuova Italia.

BERRETTA, M. (1994). «Il parlato italiano contemporaneo». In Serianni-Trifone: 239-270.

BERRUTO, G. (1987). *Sociolinguistica dell'Italiano contemporaneo*. Nuova Italia Scientifica: Milano

BERTINETTO, P.M. (1986). *Tempo, aspetto e azione nel verbo italiano*. Firenze: Accademia della Crusca.

— (1991). «Il verbo». In Renzi-Salvi: 13-161.

D'AGOSTINO, E. (1992). *Analisi del discorso*. Napoli: Loffredo.

— (1993). *Sociolinguistica computazionale*. Salerno: Ed. 10/17.

DE MAURO, T. (1963). *Storia linguistica dell'Italia unita*. Bari: Laterza.

— (1970). *Senso e significato. Studi di semantica teorica e storica*. Bari: Adriatica Ed.

— (1994) (a cura di). *Come parlano gli italiani*. Firenze: Nuova Italia.

DE MAURO, T. et alii (1993). *Lessico di frequenza dell'italiano parlato*. Milano: EtasLibri.

— (1993a). «Il Lessico di frequenza dell'italiano parlato: LIP». In AA.VV. 1993.

GAMBARARA, D. (1994). «Il passato remoto nell'italiano parlato». In De Mauro (1994): 183-194.

GENSINI, S. (1994). «Dal LIP alla didattica del parlato». In De Mauro (1994): 195-200.

GIRAUD, P. (1954). «Language et communication. Le substrat informationel de la communication». In *Bulletin de la Societé de Linguistique de Paris*, L.

JUILLAND, A.; TRAVERSA, V. (1973). *Frequency Dictionary of Italian Words*. L'Aja: Mouton.

KATERINOV, K.; BORIOSI, M.C.; SCIARONE, A.G. (1991). *Calcolo di rendimento di liste di base: italiano parlato, italiano scritto. Livello soglia. Ricerche di Linguistica Computazionale*. Perugia: Guerra.

KNEASE, T.M. (1993). *An Italian Word List from Literary Sources*. Toronto.

MIGLIORINI, B. (1943). *Der Grundlegende Wortschatz des Italienischen*. Marburg.

MOREAU, R. (1962). «Au sujet de l'utilisation de la notion de fréquence en linguistique». *Cahiers de Lexilogie*, III.

RENZI, L. (1994). «Egli - lui - il – lo». In De Mauro (1994): 247-250.

RENZI, L.; SALVI, G. (1991) (a cura di). *Grande grammatica italiana di consultazione*. Bologna: Il Mulino.

SERIANNI, L.; TRIFONE, P. (1994) (a cura di). *Storia della lingua italiana*, vol. II. Torino: Einaudi: 239-270.

SIMONE, R. (1990). *Fondamenti di linguistica*. Bari: Laterza.

SKINNER, L.H. (1935). «A Comparative Study of the Vocabularies of Forty-Five Italian Textbooks». *Modern Language Journal*, XX.

SOBRERO, A. (1993) (a cura di). *Introduzione all'italiano contemporaneo*, vol. I e II. Bari: Laterza.

STAMMERJOHANN, H. (1994). «L'italiano L2: dalla descrizione alla didattica». In De Mauro (1994): 83-87.

THOMPSON, M.E. (1927). *A Study in Italian Vocabulary Frequency*, tesi di Master. Univ. of Yowa.

VELI (1989). *Vocabolario Elettronico della Lingua Italiana. Il vocabolario del 2000.* Milano: IBM Italia.

ZAMPOLLI, A. et alii (1972). *Lessico di frequenza della lingua italiana contemporanea.* Milano: Garzanti.